

LLM Forecasting Aid (excessively preliminary)

Conversation started at 01:08:09 UTC on 2025-08-03

Session ID: 43cd4dc6-f2b5-4a56-b5d3-9be876e3fc36

User at 01:08:09

I'm answering the following question:

Will there be a novel pathogen that kills over 25 million people between 2025 and 2035 (inclusive)?

Assistant at 01:08:09

Great! Let's work through this forecasting question together. I'll help guide you through a structured approach to develop your forecast.

Step 1: Clearly defining what we're forecasting

Let me make sure we're both clear on the specifics of this question. We're forecasting whether:

- A novel pathogen (new/previously unknown disease-causing agent)
- Will kill over 25 million people
- During the time period 2025-2035 (inclusive) - so that's an 11-year window
- The deaths would need to be attributable to this single novel pathogen

Does this match your understanding of the question? Are there any aspects of the definition you'd like to clarify or discuss?

Step 2: Your initial thoughts

Now I'd like to hear your initial thoughts on this question. Please share:

- Your best current guess at a probability (as a percentage)
- The key reasoning behind your estimate
- What factors you think are most important to consider

What's your initial take on this forecasting question?

User at 01:09:44

Yep! That sounds about right.

I think the probability is 45%. COVID happened recently and caused a lot of deaths. Bird flu now seems to be a big risk, as well as Mpox. Given that a pandemic can come from many sources, I put the probability at 45%, just under a coin flip.

Assistant at 01:09:44

Thanks for sharing your initial thoughts! A 45% probability is quite high - essentially saying this is nearly as likely to happen as not. Let me help you refine this forecast.

Your reasoning touches on some important recent events and current risks. You're right that COVID-19 demonstrated that novel pathogens can cause massive casualties, and that there are multiple potential sources of future pandemics.

However, I think there's a valuable exercise that could help improve your forecast accuracy here: calculating a base rate for an appropriate reference class.

A base rate is the frequency with which similar events have occurred historically. Good forecasters often start with this "outside view" before adjusting based on specific current factors (the "inside view").

For this question, you might want to consider: How often have novel pathogens killed over 25 million people in previous decades or centuries? This could give you a useful starting point before factoring in modern considerations like global connectivity, medical advances, surveillance systems, etc.

Would you like to try calculating a base rate for an appropriate reference class, or would you prefer more guidance on how to approach this exercise?

Also, I'm curious - when you mention COVID, bird flu, and Mpox as factors pushing your probability up, have you considered what specifically makes the 2025-2035 period different from other 11-year periods in terms of pandemic risk?

User at 01:10:47

I found some data on pandemics from Our World in Data. Since 1900, it looks like the Spanish Flu, HIV/AIDS, and COVID-19 cross the 25+ million threshold. That is 3 pandemics for 130 years, so we have $3/13=0.23$ large pandemics per decade.

Does this calculation sound reasonable to you? What am I supposed to do with this result?

Assistant at 01:10:47

Great work on finding data and attempting a base rate calculation! You're on the right track with using historical data as a starting point.

Let me give you some feedback on your calculation and approach:

Your calculation: You found 3 pandemics over 130 years (1900-2030), giving you $3/13 \approx 0.23$ large pandemics per decade. For an 11-year period (2025-2035), this would suggest roughly 0.25 probability, or 25%.

Some things to consider about your reference class:

- Does your calculation account for the fact that we're looking specifically for novel pathogens? (HIV was novel, Spanish flu was novel, but was COVID-19 truly novel or a variant of known coronaviruses?)
- Are you confident that 1900-2024 is the right time period to use as your reference class, or might more recent decades be more relevant given changes in global connectivity, population density, medical capabilities, etc.?

What to do with this result: This 25% base rate gives you an "outside view" starting point. Now you can adjust up or down based on "inside view" factors specific to 2025-2035.

Your original estimate was 45% - significantly higher than this base rate. What specific factors about the 2025-2035 period make you think the risk is nearly double the historical average? Are those factors strong enough to justify such a large adjustment from the base rate?

What do you think about this approach?

User at 01:12:58

Can you say more about your question, "Are you confident that 1900-2024 is the right time period to use as your reference class?"

I put a high forecast of 45% because I think globalization increases the risks of pandemics, but I also want to make sure I'm including enough data. Would 2 pandemics (HIV and COVID-19) be enough data if I used a shorter period?

Assistant at 01:12:58

Excellent question! You've identified a key tension in base rate calculations - the trade-off between relevance and sample size.

The trade-off you're facing:

- Longer time period (1900-2024): More data points (3 pandemics), but includes eras that may be less relevant to today's world
- Shorter time period (e.g., 1980-2024): More relevant to modern conditions, but fewer data points (2 pandemics), making the estimate less statistically reliable

Some questions to help you think through this:

- When did globalization really accelerate in ways that might affect pandemic spread? 1950s? 1980s? 2000s?
- When did other relevant factors change significantly - like air travel patterns, urbanization rates, or global health surveillance systems?
- The Spanish Flu occurred in 1918-1920, in a very different world. How confident are you that it's a good predictor for 2025-2035 risks?

Alternative approaches to consider:

- You could calculate base rates for different time periods and see how they compare
- You could weight more recent pandemics more heavily in your calculation
- You could look at near-misses (SARS, MERS, Ebola outbreaks) to get a sense of how often dangerous novel pathogens emerge, even if they don't reach the 25M threshold

What's your intuition about which time period might give you the most relevant reference class for the 2025-2035 period?